Adaptive Sampling Algorithm Applied to SCADA Datalog Database Shrinkage

Mile Petkovski
and Mitko Kostov
Faculty of Technical Sciences
University St. Kliment Ohridski
7000 Bitola, Macedonia
e-mail: mile.petkovski@uklo.edu.mk
e-mail: mitko.kostov@uklo.edu.mk

Sofija Bogdanova
and Momcilo Bogdanov
Faculty of Electrical Engineering
and Information Technologies
University Ss. Cyril and Methodius
1000 Skopje, Macedonia
e-mail: sofija@feit.ukim.edu.mk

***Abstract*—In This paper an application of adaptive sampling algorithm is considered. Adaptive sampling is utilized as an efficient tool for records reduction of SCADA Datalog Database. Decreasing the number of records of Datalog File significantly increase the data-logging period. Proposed procedure is tested on commercially installed data-logging system and results are presented graphically as a resulting trend.**

***Keywords*—nonuniform sampling; discrete wavelet transform; wavelet packets; datalogging**

I. Introduction

SCADA[1] systems are used to control dispersed assets where centralized data acquisition is as important as control [4]. These systems are used in distribution systems such as water distribution and wastewater collection systems, oil and gas pipelines, electrical utility transmission and distribution systems, and rail and other public transportation systems [3], [4]. SCADA systems integrate data acquisition systems with data transmission systems and HMI[2] software to provide a centralized monitoring and control system for numerous process inputs and outputs. SCADA systems are designed to collect field information, transfer it to a central computer facility, and display the information to the operator graphically or textually, thereby allowing the operator to monitor or control an entire system from a central location in real time. Based on the sophistication and setup of the individual system, control of any individual system, operation, or task can be automatic, or it can be performed by operator commands.

SCADA systems consist of both hardware and software. Typical hardware includes an MTU[3] placed at a control center, communications equipment (e.g., radio, telephone line, cable, or satellite), and one or more geographically distributed field sites consisting of either an RTU[4], which controls actuators and/or monitors sensors. The MTU stores and processes the information from RTU inputs and outputs, while the RTU controls the local process. The communications hardware allows the transfer of information and data back and forth between the MTU and the RTUs. The software is programmed to tell the system what and when to monitor, what parameter ranges are acceptable, and what response to initiate when parameters go outside acceptable values.

The data-logging system is a centralized database for logging all measured data. Information stored in this database can be accessed to support various analysis, from statistical process control to enterprise level planning. The data are recorded periodically and presented graphically by trend charts. Trends can show real time measured data or pool data-logged values. The data-logging database file increased its size according to signal sampling rate.

In this paper an algorithm for adaptive sampling is utilized to reduce the number of database records, but to keep trend charts readable. The paper is organized as follows. Section II presents the theoretical fundamentals for applied algorithm. A brief description of the targeted data-logging database system is given in Section III. The proposed algorithm implementation is presented in section IV and comments and conclusions in section V.

II. Theoretical Background

A. Discrete Haar Wavelet Transform

In series expansion of discrete-time function f using wavelets [5]

$$f(t) = \sum_{j=1}^{J} \sum_{k=1}^{2^{-j}M} d_{jk}\psi_{jk}(t) + \sum_{k=1}^{2^{-J}M} a_{Jk}\varphi_{Jk}(t), \qquad (1)$$

ψ_{jk} and ϕ_{jk} denote wavelet and scaling function, respectively, the indexes j and k are for dilatation and translation, and a_{Jk} and d_{jk} are approximation and detail coefficients.

Wavelet decompositions and multiresolution concepts are closely related to filter bank theory. For this reason, it is helpful to view the scaling and wavelet function as a low pass and high pass filter, H_0 and H_1, respectively. The wavelet transform is applied to low pass results (approximations) only as it is illustrated in Fig. 1.

[1] Supervisory Control And Data Acquisition
[2] Humane Machine Interface
[3] Master Terminal Unit
[4] Remote Terminal Unit

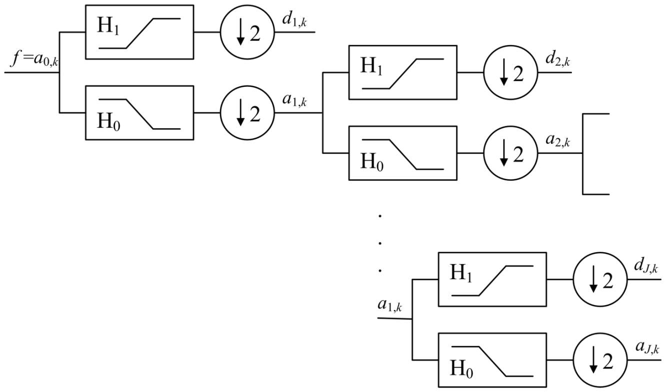

Figure. 1. Discrete wavelet transform tree.

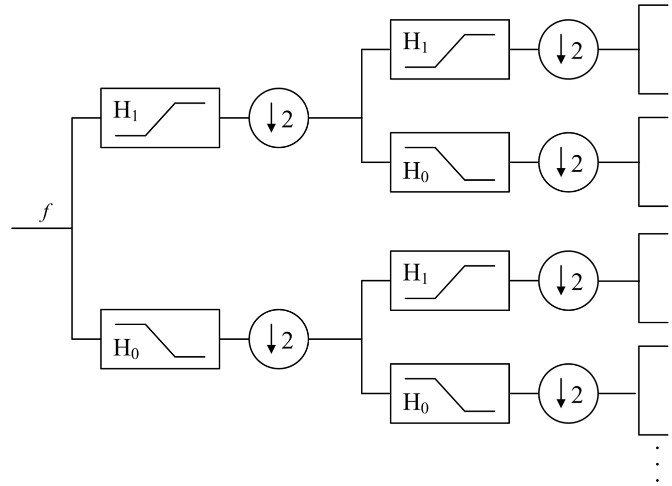

Figure. 2. Wavelet packet decomposition.

The Haar wavelet's mother wavelet function $\psi(t)$ and scaling function $\phi(t)$ can be described as

$$\psi(t) = \begin{cases} 1 & 0 \le t < 1/2 \\ -1 & 1/2 \le t < 1 \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$

$$\phi(t) = \begin{cases} 1 & 0 \le t < 1 \\ 0 & \text{otherwise} \end{cases}, \tag{3}$$

respectively. The corresponding low pass and high pass filters are $[1,1]/\sqrt{2}$ and $[1,-1]/\sqrt{2}$, which means that the haar wavelet decomposes a signal into two sub-signals: running average and running difference.

B. Wavelet Packets

WPT[5] is an extension of Discrete Wavelet Transform and can be obtained by a generalization of the fast pyramidal algorithm. The transform is applied to both low pass results (approximations) and high pass results (details), i.e. in wavelet packet analysis the details as well as the approximations are splited. The complete binary tree is produced as presented in Fig. 2.

The WPT transform and provides a more flexible tool for the time-scale analysis of data.

Wavelet packet decomposition gives a lot of bases from which you can look for the best representation with respect to design objectives.

C. Adaptive Sampling Algorithm

The purpose of adaptive sampling algorithm [2] is to obtain a subset of non-uniformly distributed samples $\mathbf{y}= [y_0, y_1 \dots y_{P-1}]^T$, from given N uniformly distributed samples of a discrete signal $\mathbf{x}= [x_0, x_1 \dots x_{N-1}]^T$, where $P < N$. At the same time it is necessary to keep enough information for satisfactory signal reconstruction applying the suitable reconstruction algorithm which require a small amount of resources and computational effort. The sampling rate estimation procedure is based on the second stage wavelet packet coefficients. More precisely, coefficients from the second step of wavelet packet transform are used to estimate the sampling rate of a single interval. To illustrate this we give an example. A uniformly sampled signal of 64 samples is shown in Fig. 3.

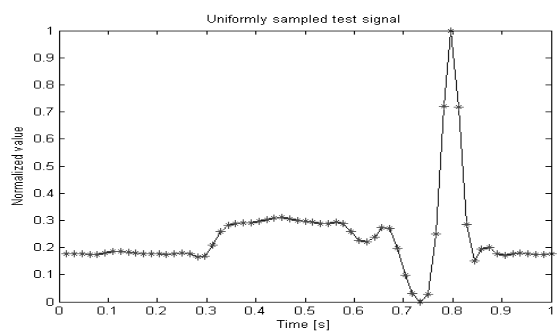

Figure. 3. Uniformly sampled test signal.

The absolute values of second stage wavelet packet transform coefficient represent a criteria for sampling rate for adaptive non uniform sampling, depicted on Fig. 4.

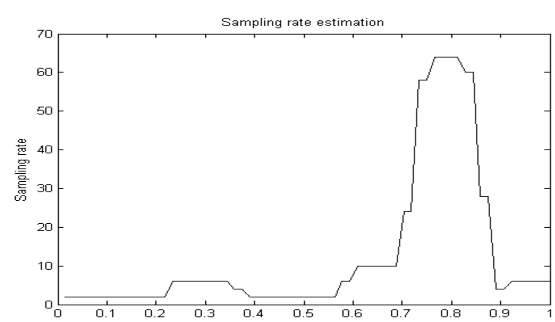

Figure. 4. Estimated sampling rate of the test signal.

[5]Wavelet Packet Transform

Faster sampling rates correspond to larger absolute values and vice versa, smaller absolute values correspond to slower sampling rates. Adaptive sampled signal according to mentioned criteria is presented on the Fig. 5.

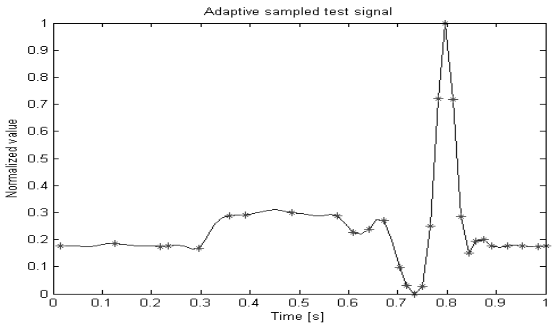

Figure. 5. Adaptive non-uniformly sampled signal.

III. SCADA Data-Logging Database Shrinkage

This section describes the basic idea of data-logging database shrinkage for FactoryTalk SCADA [1]. Motivation for database records reduction is to increase the data-logging historical period in the case of limited storage resources. The diagram in Fig. 6 shows the overall functional block diagram of the data-logging system implemented in supervisory control applications.

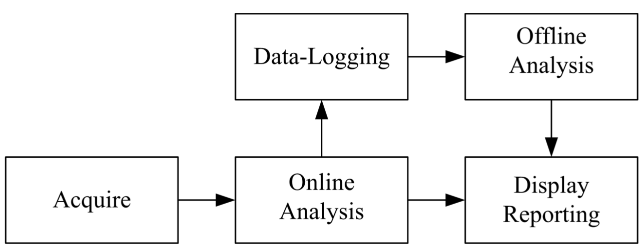

Figure. 6. Data-logging block diagram of SCADA system.

A. Trending

A trend is a visual representation, or chart, of current or historical tag values. A trend provides operators with a way of tracking plant activity as it is happening. The trend object displays real-time data and historical data from the data logs. Pens on the run-time chart represent data from the tags and expressions that you add to the trend object. The trend object provides extensive, flexible run-time control. There are two different types of Trending

- Historical Trending is when a trend polls data from a data log that is previously configured. This will allow the user to browse through a timeline to look at the data over a given period of time.
- A Runtime trend displays data trends directly from the processor. The trend will start trending its runtime data from the time it's first loaded on the display.

B. Data Logging

Data logging and recording is a very common measurement application. In its most basic form, data logging is the measurement and recording of physical or electrical parameters over a period of time. Defining a data log model in SCADA application it is specified when to collect them, and where to store them. To have a permanent record of data values, log them to the data log file on disk. The values can also be logged to an ODBC [6]-compliant database.

C. The Structure of Datalog File

The datalogging system stores measured data in a Microsoft Access database. The database contains three tables: FloatTable, StringTable and TagTable. The table TagTable has the following schema TagTable(TagName: Text, TagIndex: Number, TagType: Number, TagDataType: Number). It contains data for different signals identified by the attribute TagIndex which is the primary key. The attribute TagName is used for names/descriptions of the signals.

The table FloatTable has the following schema FloatTable(DateAndTime: Date/Time, Millitm: Number, TagIndex: Number, Val: Number, Status: Text, Marker: Text). The attributes DateAndTime and Millitm form the primary key, while the attribute TagIndex is a foreign key linked to the primary key of the table TagTable. The attribute Val contains the measured values of the signals identified by the attribute TagIndex in the moments DateAndTime.

D. The Shrinkage Procedure Description

The proposed algorithm for adaptive non-uniformly sampling of signals is utilized in the procedure for data-logging database records reduction. At the beginning the records associated to the chosen tag and specified period are imported from the FloatTable. The records are stored to the auxiliary table. After finishing the sampling rate estimation procedure the table is reorganized discarding some samples in the areas of lower sampling rate and keeping more samples in the areas with higher sampling rate. The table content represents an adaptive sampled original signal as a sequence of non-uniformly distributed samples. At the end, the sequence of uniformly distributed samples in the FloatTable for chosen tag and specified period is replaces by the contents of the auxiliary table.

IV. Implementation

Proposed dataloging database shrinkage algorithm is applied to the datalog file for storing the induction motor current measured values. Uniformly distributed sequence of samples stored in the datalog file is presented by the trend chart depicted in the Fig. 7.

Data-logging file records are imported by the procedure purposed for sampling rate estimation and adaptive sampling.

The first stage of adaptive sampling algorithm is sampling rate estimation described in details in the previous text. As

[6]Open Data-Base Conectivity

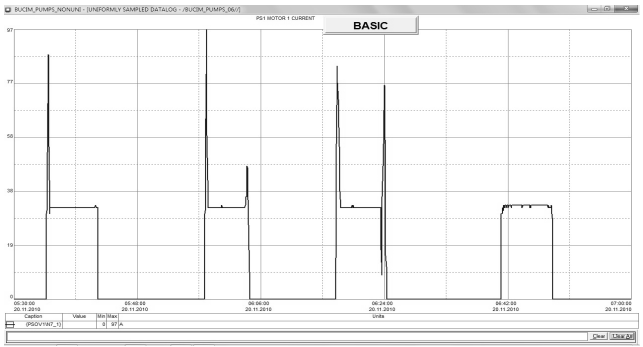

Figure. 7. Trend of uniformly sampled signal.

a tool for sampling rate estimation is used a wavelet packet decomposition tree.

The resulting coefficients of wavelet packet transform are presented on Fig. 8.

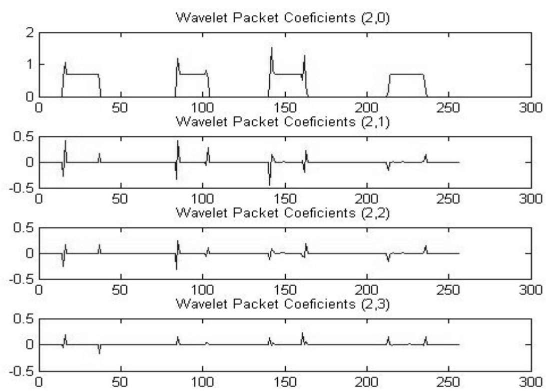

Figure. 8. Wavelet packet decomposition coefficients.

Fig. 9 presents the estimated sampling rate for adaptive sampling. At the areas of constant value or linear signal changes the sampling rate is considerable less.

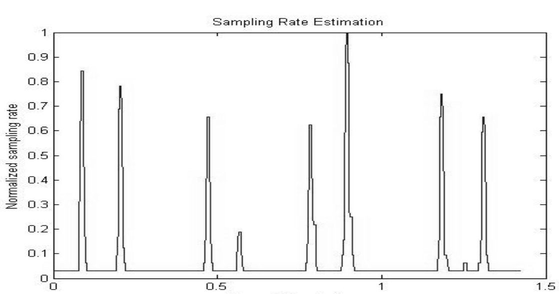

Figure. 9. Estimated sampling rate of measured signal.

Otherwise in the areas where signal emerges significant changes for the short time the sampling rate is closer to the Niquyst rate.

In the experiment, 32 different sampling periods are used. The shortest sampling period corresponds to the sampling period of the original uniformly sampled signal. The lowest sampling frequency corresponds to the period 32 times longer then the sampling period of original signal. According to the results presented in Fig. 9, the original uniformly sampled signal is resampled, which yields the non-uniformly adaptive sampled signal.

Resulted adaptive nonuniformly sampled signal is exported to the database and coresponding trend is presented in the Fig. 10.

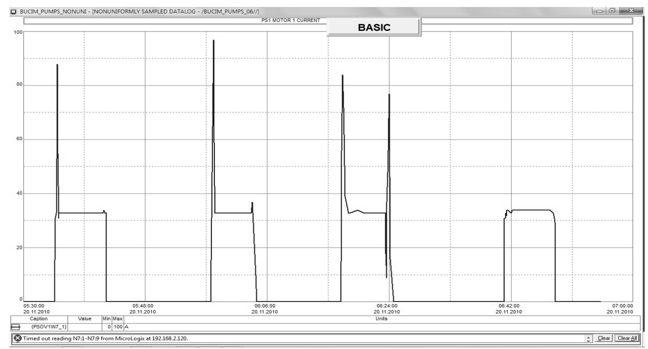

Figure. 10. Adaptive sampled measured current trend.

In Table 1 a numerical results obtained by the implementation procedure are presented. The number of records and occupied storage in the case of uniformly and non-uniformly sampled signal are considered.

TABLE I
NUMERICAL RESULTS

*	**Uniformly Sampled**	**Adaptive Sampled**
Records	1024	176
Storage [kB]	1488	196

V. CONCLUSION

We have presented a successful application of the proposed adaptive sampling algorithm in SCADA systems in order to reduce the number of datalog database records keeping trend charts readable. Results confirm that records reduction is significant and datalog database occupied storage is considerably less. We believe that this algorithm can be useful tool in all applications where the reduction of data is important.

REFERENCES

[1] Rockwell Automation, FactoryTalk View Site Edition, User's Guide, Publication VIEWSE-UM006E-EN-EJune 2009
[2] Petkovski Mile, Bogdanova Sofija and Bogdanov Momcilo, A Simple Adaptive Sampling Algorithm, XIV Telecomunications Forum - TELFOR 2006, Belgrade, Serbia, November 21 - 23, 2006.
[3] Keith Stouffer, Joe Falco, Karen Kent, Guide to Supervisory Control and Data Acquisition (SCADA) and Industrial Control Systems Security, NIST Special Publication 800-82, September 2006
[4] David Bailey, Edwin Wright, Practical SCADA for Industry, Newnes - Elsevier, Copyright (c) 2003, IDC Technologies
[5] G. Strang, T. Nguyen, Wavelets and Filter Banks, Wellesley-Cambridge Press, Wellesley MA, 1995.